International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN 2319-3972
Vol. 2, Issue 3, July 2013, 1-6

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

REGRESSION STUDY FOR BREAST CANCER IN AL-SADER MEDICAL CITY

HADEEL SALIM ALKUTUBI

Informatics Center for Research & Rehabilitation, Kufa University, Iraq

ABSTRACT

In this study, Breast Cancer in Al-Sader Medical City was introduced. Regression analysis is used to analyzed the data to get the mathematical model and the effect between all variables. In the results, there exist a positive relationship between Y (tumor levels) and X1 (education) in the years 2005, 2008, 2009, 2010, and 2011. But a negative relationship in the years 2006, 2007, and 2012.

KEYWORDS: Linear Regression, Correlation Coefficient, Breast Cancer, SPSS Program

INTRODUCTION

Linear multiple regression analysis extends simple linear regression analysis by considering two or more independent variables. In the case of three independent variables, denoted as X_1, X_2, X_3, we use the estimated multiple regression equation $Y_C = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ here, Y_C, with the subscript C for "computed", denotes values for Y calculated from the estimated regression equation.

This is analogous to the term $\overline{Y_x}$ we used before, but Y_C is used in dealing with several independent variables because it is too cumbersome to place X_1, X_2, X_3 in subscripts. With three independent variables and one dependent variable, a total of four variables must be considered. The sample data will consist of four values for each sample unit observed.[Lawrence].

We taken 900 case from Al-Sader medical city in Al-Najaf in Iraq for the period 2005-2009. This data contain breast cancer divided between Levels of Tumor, Occupation, Marital Status and Education. The main aim of this study are presentation and description all cases of breast cancer and find the significant difference between all years and groups depend on all variables levels of tumor, occupation, marital status and education using SPSS program to get regression models summary, analysis of variance for all variables, and regression coefficients.

MATERIALS AND METHODS

Multiple regression may use more than two independent variables. For example three independent variables, denoted by X_3, would provide the regression equation $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. In solving for the regression coefficient, additional product sums involving X_3 must be computed and four normal equations must be solved simultaneously. Ordinarily, the necessary calculations are made with a digital computer rather than by hand.

As in multiple regression with $X_1$and X_2, by including X_3 we my compute the standard error of the estimate about the regression hyper plane. (including Y, there will now be four variables, so the regression surface cannot be graphed in only three dimensions). The standard error, denoted by the symbol $S_{Y.123,}$ is calculated in the same way we calculate $S_{Y.12,}$, but with an extra term involving β_3 in the numerator; the denominator will be n-4= 10-4 = 6.

The number of degrees of freedom will also be n-4 , and we may calculate the prediction intervals as we did before with this reduced value.

When there are m total variables (one dependent and m-1 independent), the number of degrees of freedom will be n-m, which is used as the denominator for computing the standard error and in finding t_α when constructing prediction intervals. [Lawrence].

To find regression analysis, we used SPSS program in the next section to get regression models summary, analysis of variance for all variables, and regression coefficients.

RESULTS

In this section, we used SPSS program to get regression models summary (Table 1), analysis of variance for all variables (Table 2), and regression coefficients (Table 3) as follows.

Table 1: Regression Model Summary for Education, Occupation, Marital Status and Tumor in 2005-2012

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate
2005	0.353[a]	0.125	-0.167	1.40631
2006	0.113[a]	0.013	-0.001	1.06362
2007	0.129[a]	0.017	0.002	0.82895
2008	0.139[a]	0.019	0.008	0.99985
2009	0.164[a]	0.027	0.013	0.93343
2010	0.027[a]	0.001	-0.008	1.02533
2011	0.096[a]	0.009	-0.003	1.00707
2012	0.171[a]	0.029	-0.054	1.17459

a. Predictors: (constant), Education, Occupation, and Marital Status

Table 2: Analysis of Variance for Education, Occupation, Marital Status and Tumor in 2005-2012

Years	Model	Sum of Squares	DF	Mean Square	F	Sig.
2005	Regression	1.689	2	0.845	0.427	0.671[a]
	Residual	11.866	6	1.978		
	Total	**13.556**	**8**			
2006	Regression	3.126	3	1.042	0.921	0.431
	Residual	239.832	212	1.131		
	Total	**242.958**	**215**			
2007	Regression	2.411	3	0.804	1.169	0.322[a]
	Residual	142.243	207	0.687		
	Total	**144.654**	**210**			
2008	Regression	4.927	3	1.642	1.643	0.180[a]
	Residual	249.927	250	1.000		
	Total	**254.854**	**253**			
2009	Regression	4.936	3	1.645	1.888	0.133[a]
	Residual	179.488	206	0.871		
	Total	**184.424**	**209**			
2010	Regression	0.260	3	0.087	0.082	0.970[a]
	Residual	359.547	342	1.051		
	Total	**359.806**	**345**			
2011	Regression	2.335	3	0.778	0.767	0.513[a]
	Residual	251.519	248	1.014		
	Total	**253.853**	**251**			
2012	Regression	1.455	3	0.485	0.352	0.788[a]
	Residual	48.288	35	1.380		
	Total	**49.744**	**38**			

a. Predictors: (constant), Education, Occupation, and Marital Status
b. Dependent variable: Tumor

Table 3: Regression Coefficients for Education, Occupation, Marital Status and Tumor in 2005-2012

Years	Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
2005	Constant	3.104	1.094	0.333	2.837	0.030
	Education	0.264	0.304		0.868	0.419
	Occupation	-0.452	1.133	-0.153	-0.399	0.704
	Marital Status					
2006	Constant	3.428	0.207	-0.078	16.551	0.000
	Education	-0.059	0.057		-1.042	0.299
	Occupation	0.697	0.483	0.108	1.443	0.151
	Marital Status	-0.144	0.199	-0.050	-0.722	0.471
2007	Constant	3.651	0.137	-0.081	26.580	0.000
	Education	-0.047	0.040		-1.176	0.241
	Occupation	-0.227	0.216	-0.073	-1.050	0.295
	Marital Status	-0.101	0.123	-0.057	-0.818	0.414
2008	Constant	3.005	0.152	0.107	19.769	0.000
	Education	0.076	0.053		1.424	0.156
	Occupation	-0.192	0.227	-0.064	-0.844	0.399
	Marital Status	0.204	0.114	0.113	1.791	0.075
2009	Constant	2.888	0.167	0.124	17.346	0.000
	Education	0.082	0.046		1.781	0.076
	Occupation	0.309	0.178	0.124	1.741	0.083
	Marital Status	0.103	0.140	0.053	0.733	0.464
2010	Constant	3.223	0.129	0.013	25.055	0.000
	Education	0.010	0.043		0.231	0.818
	Occupation	0.026	0.266	0.005	0.099	0.921
	Marital Status	0.044	0.107	0.022	0.408	0.683
2011	Constant	3.172	0.173	0.082	18.330	0.000
	Education	0.065	0.051		1.290	0.198
	Occupation	-0.317	0.386	-0.052	-0.821	0.413
	Marital Status	0.028	0.170	0.011	0.166	0.868
2012	Constant	3.555	0.616	-0.050	5.774	0.000
	Education	-0.036	0.123		-0.296	0.769
	Occupation	-0.650	0.862	-0.127	-0.755	0.456
	Marital Status	-0.313	0.622	-0.084	-0.503	0.618

a. Dependent variable: Tumor

DISCUSSIONS

In Table 1, we can see from 2005 the R^2 = 1.3 %, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is week. In 2006 the $R^2 = 1.3$ %, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is week. Also in 2007 the $R^2 = 1.7$ % , this means that Education effect , Occupation effect, and Marital Status effect of tumor level is week.

From 2008, we can see the $R^2 = 1.9$, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is week. And from 2009, we get $R^2 = 2.7$, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is week. But in 2010 and 2011, the $R^2 = 0.1\ and\ R^2 = 0.9$ %, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is very week. In 2012 the $R^2 = 2.9\%$, this means that Education effect, Occupation effect, and Marital Status effect of tumor level is week.

From Table 2, we get analysis of variance for education, occupation, marital status and tumor levels in the period 2005-2012. We can get there is not significant different between all variables (education, occupation, marital status and

tumor) from 2005 to 2012. From Table 3, we see the regression line in 2005 is $\hat{Y} = 3.104 + 0.264X_1 - 0.452X_3$, this means that, there is a positive relationship between Y (Tumor) and X_1 (education), but a negative relationship between Y (Tumor) and X_3 (marital status) . In 2006, the regression line is $\hat{Y} = 3.428 - 0.059X_1 + 0.697X_2 - 0.144X_3$, this means that, there is a negative relationship between Y (Tumor) and X_1 (education), and a positive relationship between Y (tumor) and X_2 (occupation) but a negative relationship between Y (Tumor) and X_3 (marital status). In 2007 the regression line is $\hat{Y} = 3.651 - 0.047X_1 - 0.227X_2 - 0.101X_3$, this means that, there is a negative relationship between Y (Tumor) and X_1 (education), and a negative relationship between Y (tumor) and X_2 (occupation) but a negative relationship between Y (Tumor) and X_3 (marital status).

In 2008, the regression line is $\hat{Y} = 3.005 + 0.076X_1 - 0.192X_2 + 0.204X_3$, this means that, there is a positive relationship between Y (Tumor) and X_1 (education), and a negative relationship between Y (tumor) and X_2 (occupation) but a positive relationship between Y (Tumor) and X_3 (marital status). In 2009, the regression line is $\hat{Y} = 2.888 + 0.082X_1 + 0.309X_2 + 0.103X_3$, this means that, there is a positive relationship between Y (Tumor) and X_1 (education), and a positive relationship between Y (tumor) and X_2 (occupation) and positive relationship between Y (Tumor) and X_3 (marital status). In 2010, the regression line is $\hat{Y} = 3.223 + 0.010X_1 + 0.026X_2 + 0.044X_3$, this means that, there is a positive relationship between Y (Tumor) and X_1 (education), and a positive relationship between Y (tumor) and X_2 (occupation) and positive relationship between Y (Tumor) and X_3 (marital status).

In 2011, the regression line is $\hat{Y} = 3.172 + 0.065X_1 - 0.317X_2 + 0.028X_3$, this means that, there is a positive relationship between Y (Tumor) and X_1 (education), and a negative relationship between Y (tumor) and X_2 (occupation) but a positive relationship between Y (Tumor) and X_3 (marital status). Finally in 2012, the regression line is $\hat{Y} = 3.555 - 0.036X_1 - 0.650X_2 - 0.313X_3$, this means that, there is a negative relationship between Y (Tumor) and X_1 (education), and a negative relationship between Y (tumor) and X_2 (occupation) also a negative relationship between Y (Tumor) and X_3 (marital status).

CONCLUSIONS

- Education effect , Occupation effect , and Marital Status effect of tumor level is a week in the years 2005, 2006, 2007, 2008, 2009, and 2012. But Education effect , Occupation effect , and Marital Status effect of tumor level is a very week in the years 2010 and 2011.
- There is not significant different between all variables (education, occupation, marital status and tumor) from 2005 to 2012.
- There exist a positive relationship between Y (tumor levels) and X_1 (education) in the years 2005, 2008, 2009, 2010, and 2011. But a negative relationship in the years 2006, 2007, and 2012.
- There exist a positive relationship between Y (tumor levels) and X_2 (occupation) in the years 2006, 2009, and 2010. But a negative relationship in the years 2007, 2008, 2011, and 2012.
- There exist a positive relationship between Y (tumor levels) and X_3 (marital status) in the years 2008, 2009, 2010, and 2011. But a negative relationship in the years 2005, 2006, 2007, and 2012.

REFERENCES

1. Alkutubi H. S., 2009. On regression modeling of human immunodeficiency virus patients. American journal of applied sciences. Science publication. 6(8):1580- 1585.

2. Alkutubi H. S., Ibrahim N. A. and Yaseen N. K., 2009. On statistical analysis of cancer tumors in tikrit hospital. European journal of scientific research. Euro journals publishing, Inc. 35(1):106-120.
3. Alkutubi H.S, 2005. Evaluation of cancer disease for the period (1995-2005) in Tikrit teaching hospital, Tikrit journal of pharmaceutical science,vol.1,no.2.
4. Bland Martin, 2000. An introduction to medical statistics, 3th ed. Oxford University Press.
5. Lawrence L. L. 1975. Statistics meaning and method. Harcourt brace Jovan.